# Analysis and End of Covid-19 in India

Ulhas J Dixit[1] and Priyesh D Tiwari[2]
Department of Statistics, University of Mumbai

### Abstract

We have collected data on Covid-19 from Arogya Set and www.kaggle.com for India. In this case we have analysed the data on covd-19 by using regression and time series. Using these techniques, we have predicted the total positive cases, cumulative positive cases of covid-19 up to 25th July and also predicted the date when the covid-19 pandemic ends.

**Key words and phrases**: Covid-19, regression, time series, end of pandemic, India.

### 1.Introduction

COVID-19 pandemic in India is part of the worldwide pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first case of COVID-19 in India, which originated from China, was reported on 30 January 2020

On 22 March, India observed a 14-hour voluntary public curfew at the instance of the prime minister Narendra Modi. It was followed by mandatory lockdowns in COVID-19 hotspots and all major cities. Further, on 24 March, the Prime Minister ordered a nationwide lockdown for 21 days, affecting the entire 1.3 billion population of India. On 14 April, the PM extended the nationwide lockdown till 3 May which was followed by two-week extensions starting 3rd and 17th May with substantial relaxations. On 1st June the Government started unlocking the country (barring containment zones) in three steps.

As of 19 June 2020, the Ministry of Health and Family Welfare (MoHFW) has confirmed a total of 380,532 cases, 204,711 recoveries (including 1 migration) and 12,573 deaths in the country.

The prevention of spread of covid-19 has proven to be a big challenge. Moreover, the testing of suspects and treatment of infected has been another big challenge. The government/administration must be prepared with the resources required for the testing and the treatment and if they go scarce it might prove to be very troublesome in the light of surge of virus in the near future. This strenuous task can be diluted and dealt easily if the number of positive cases and can be anticipated in advance so that we are well prepared to confront this pandemic.

### About the Data

We have received the data from Aarogya-Setu and www.kaggle.com . The data available is

**Cumulative Positive**: cumulative number of positive cases

**Daily Positive**: number of positive cases found daily

**Active Cases:** the number of covid-19 active cases.

**Recovered Cases:** total number of recovered cases until a given day

For Cumulative and Daily Positive Cases, the data is considered from 2nd March 2020 to 24th June 2020 and for active and recovered cases the data is available from 4th April 2020 to 16th July 2020.

### 2. Graphical Presentation

We start with first visualizing the data through graph. The Daily +ve cases are plotted against the date in the Figure 1.

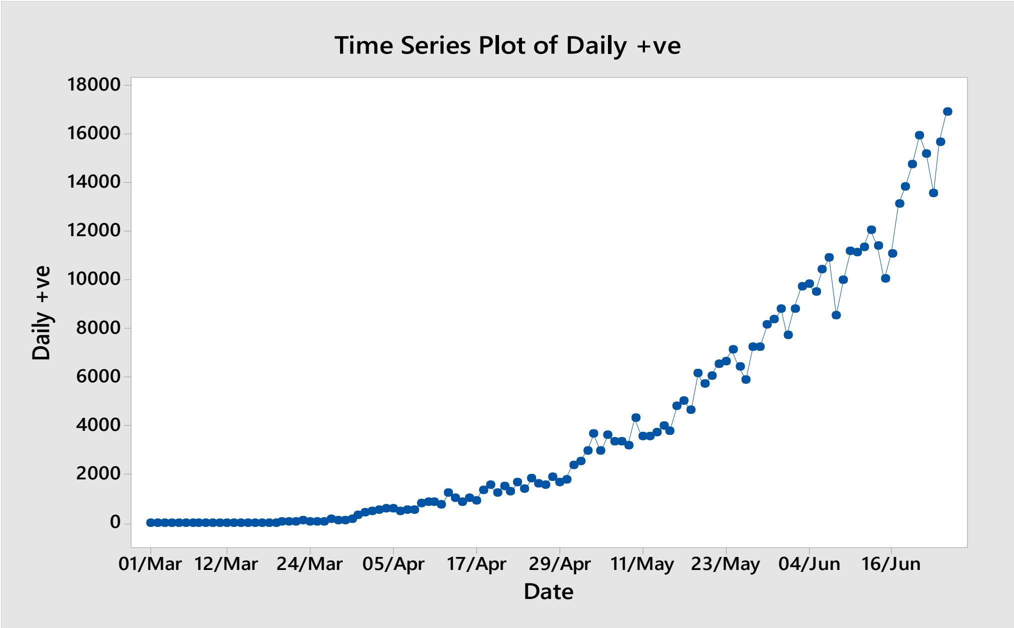


*Figure 1: India*

It is natural to think to model the data using regression analysis, where Daily +ve or Cumulative +ve can be taken as the response and the variable Day as the predictor if the purpose is the long-term prediction. However, the problem is, the predictor in the regression analysis is a controllable variable but there are other factors such as lockdown, invention of the drug/vaccine etc. However, despite these uncontrollable factors playing role if one can model the relationship between Cumulative or Daily +ve and Day and if it satisfies the theoretical assumptions and parameters satisfy our requirements then these models can be used for predicting the response for long term.

The prediction will be valid only if the uncontrollable factors, which determine the rate of change in the response over the time, remain roughly the same. For instance, if the government takes some drastic measures to control the spread of the virus or the vaccine of the disease is invented or for any reason in the near future then these incidents are very likely to affect the number of positive cases detected daily.

We fit the regression model for Daily +ve against Day for India.

## 3. Regression Analysis: Daily +ve versus Day

We used Minitab to assist our statistical analysis. We start with fitting the regression model for India with Daily +ve as response variable and Day as predictor. We have already seen that the relation between Daily +ve and Day is not linear but exponential. Therefore, we use Box-Cox transformation to transform the response variable. Hence, the model we fit is

$$\textbf{(Daily+ve)}^{\lambda} = \textbf{a + b Day}$$

Where λ is the parameter of the Box-Cox transformation which is to be determined. The output produced by Minitab showed that the intercept 'a' was insignificant. Therefore, we fit a model without intercept. The optimum value of λ produced by Minitab was 0.289171 which is very close to 0.3. The 95% confidence interval for λ was (0.27667, 0.30167) which contains 0.3. Therefore, we decided to

transform the response with $\lambda = 0.3 = 3/10$ for the sake of simplicity of the model. The regression model fit is

$$\textbf{(Daily +ve)}^{0.3} = \textbf{(0.162337 Day)} \tag{1}$$

Or

$$\textbf{Daily +ve} = \textbf{(0.162337 Day)}^{10/3}$$

The ANOVA table of the model is produced in the Table 1. The p-value for regression model is too small, indicating that the model fit is overall significant.

*Table 1: Analysis of Variance for Transformed Response*

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| **Regression** | 1 | 13889.4 | 13889.4 | 43628.38 | 0.000 |
| **Day** | 1 | 13889.4 | 13889.4 | 43628.38 | 0.000 |
| **Error** | 114 | 36.3 | 0.3 | | |
| **Total** | 115 | 13925.7 | | | |

The Table 2 produces $R^2$ and $R^2$-predicted which are 99.74% and 99.73% respectively indicating that the model is very well fit. Technically one can say that about 99.74% of the variations in Daily +ve is explained by the Day.

*Table 2: Model Summary for Transformed Response*

| R-sq | R-sq(pred) |
|---|---|
| 99.74% | 99.73% |

As the model is fit without intercept and in ANOVA we have already seen that the model is significant, there is no meaning of testing the regression coefficient again.

To assess the **validity of the assumptions** of regression model the residual plot is produced in the Figure 2.

(a) The residuals are satisfactorily normal and it can be seen in the NPP at the top-left corner and histogram at the bottom-left corner of the Figure 2.
(b) The variance of the residual is almost constant as shown by the graph of residual vs fits in the top-right corner of the residual plot.
(c) The Durbin-Watson Statistic for Transformed Response is 0.7321 which indicates there is positive autocorrelation among the residuals which can be observed even in the graph in the bottom-right corner of the residual plot produced in Figure 2.

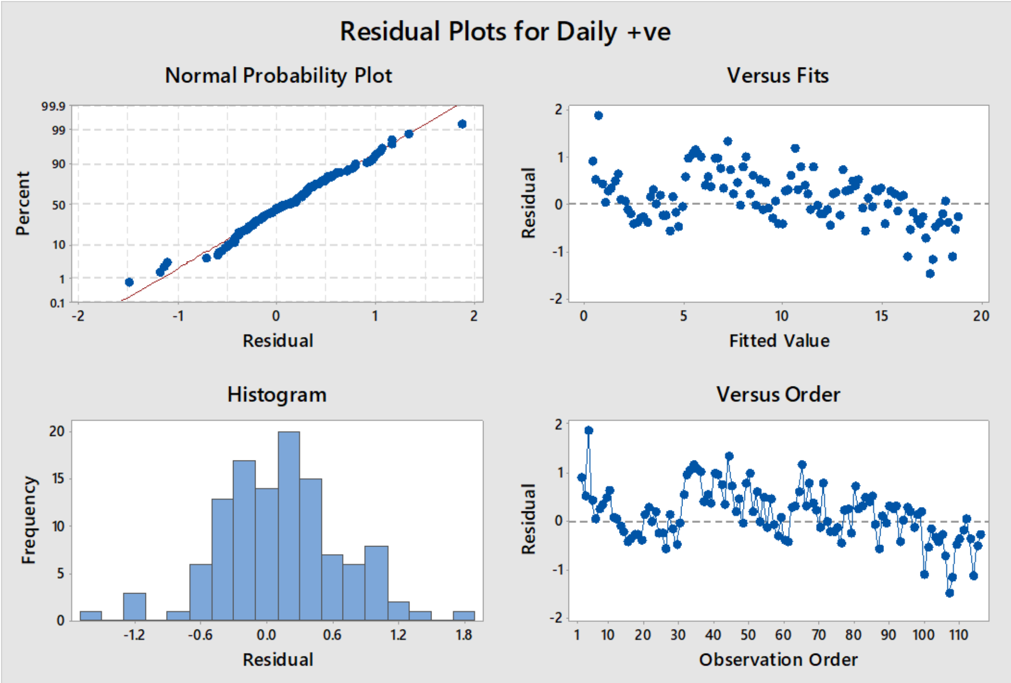


*Figure 2: Residual plot*

In such situation one can model the residuals using time series modelling to assist in predictions for short term but our purpose is long term prediction. Besides, the $R^2$ is 99.67% which almost 100%. Therefore, we believe that the contribution made by the modelled errors in prediction will be negligible. Therefore, we do not go for time series modelling of the residuals.

Further, we produce the graph of the observed and fitted Daily +ve in the Figure 3 and a table of next 30 days prediction of Daily +ve and hence Cumulative +ve with 95% prediction interval in Table 3.

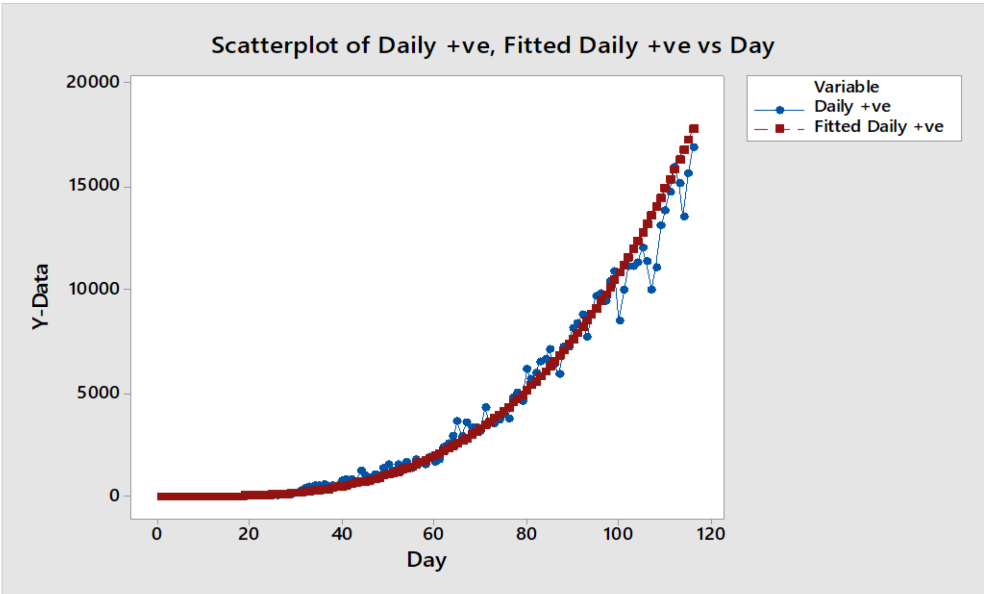


*Figure 3: observed and fitted Daily +ve*

Table 3: Predicted Daily and Cumulative +ve with Prediction limits for India

| Date | Observ-ed Daily +ve | Predicted Daily +ve | LPL of Of Daily +ve | UPL Daily +ve | Observed Cumula-tive +ve | Predicted Cumulative +ve | LPL of Cumula -tive +ve | UPL of Cumula -tive +ve |
|---|---|---|---|---|---|---|---|---|
| 1 July | 19429 | 21598 | 17774 | 25963 | 605221 | 612377 | 587156 | 641261 |
| 2 July | 21947 | 22189 | 18289 | 26636 | 627168 | 634566 | 605445 | 667897 |
| 3 July | 22718 | 22791 | 18815 | 27321 | 649886 | 657357 | 624260 | 695218 |
| 4 July | 24018 | 23404 | 19350 | 28018 | 673904 | 680761 | 643610 | 723236 |
| 5 July | 23942 | 24029 | 19897 | 28727 | 697846 | 704790 | 663507 | 751963 |
| 6 July | 22500 | 24666 | 20454 | 29450 | 720346 | 729456 | 683961 | 781413 |
| 7 July | 23145 | 25314 | 21022 | 30184 | 743491 | 754770 | 704983 | 811597 |
| 8 July | 25561 | 25974 | 21601 | 30932 | 769052 | 780744 | 726584 | 842529 |
| 9 July | 25790 | 26646 | 22191 | 31692 | 794842 | 807390 | 748775 | 874221 |
| 10 July | 27762 | 27330 | 22792 | 32465 | 822604 | 834720 | 771567 | 906686 |
| 11 July | 27757 | 28026 | 23404 | 33252 | 850361 | 862746 | 794971 | 939938 |
| 12 July | 29106 | 28735 | 24028 | 34052 | 879467 | 891481 | 818999 | 973990 |
| 13 July | 28178 | 29456 | 24663 | 34865 | 907645 | 920937 | 843662 | 1008855 |
| 14 July | 29917 | 30189 | 25310 | 35691 | 937562 | 951126 | 868972 | 1044546 |
| 15 July | 32607 | 30936 | 25969 | 36532 | 970169 | 982062 | 894941 | 1081078 |
| 16 July | 35418 | 31695 | 26640 | 37386 | 1005587 | 1013757 | 921581 | 1118464 |
| 17 July | 34824 | 32467 | 27323 | 38253 | 1040411 | 1046224 | 948904 | 1156717 |
| 18 July | 37411 | 33252 | 28018 | 39135 | 1077822 | 1079476 | 976922 | 1195852 |
| 19 July | 40235 | 34050 | 28725 | 40031 | 1118057 | 1113526 | 1005647 | 1235883 |
| 20 July | 36806 | 34862 | 29444 | 40942 | 1154863 | 1148388 | 1035091 | 1276825 |
| 21 July | 39170 | 35687 | 30176 | 41866 | 1194033 | 1184075 | 1065267 | 1318691 |
| 22 July | 45601 | 36526 | 30921 | 42806 | 1239634 | 1220601 | 1096188 | 1361497 |
| 23 July | 48443 | 37378 | 31679 | 43760 | 1288077 | 1257979 | 1127867 | 1405257 |
| 24 July |  | 38244 | 32449 | 44728 |  | 1296223 | 1160316 | 1449985 |
| 25 July |  | 39125 | 33232 | 45712 |  | 1335348 | 1193548 | 1495697 |
| 26 July |  | 40019 | 34029 | 46711 |  | 1375367 | 1227577 | 1542408 |
| 27 July |  | 40927 | 34838 | 47724 |  | 1416294 | 1262415 | 1590132 |
| 28 July |  | 41850 | 35662 | 48754 |  | 1458144 | 1298077 | 1638886 |
| 29 July |  | 42787 | 36498 | 49798 |  | 1500931 | 1334575 | 1688684 |
| 30 July |  | 43739 | 37348 | 50858 |  | 1544670 | 1371923 | 1739542 |
| 31 July |  | 44706 | 38212 | 51934 |  | 1589376 | 1410135 | 1791476 |
| 1 Aug |  | 45687 | 39090 | 53026 |  | 1635063 | 1449225 | 1844502 |
| 2 Aug |  | 46684 | 39982 | 54134 |  | 1681747 | 1489207 | 1898636 |
| 3 Aug |  | 47695 | 40888 | 55258 |  | 1729442 | 1530095 | 1953894 |
| 4 Aug |  | 48722 | 41809 | 56398 |  | 1778164 | 1571904 | 2010292 |
| 5 Aug |  | 49764 | 42743 | 57554 |  | 1827928 | 1614647 | 2067846 |
| 6 Aug |  | 50822 | 43693 | 58727 |  | 1878750 | 1658340 | 2126573 |
| 7 Aug |  | 51895 | 44657 | 59917 |  | 1930645 | 1702997 | 2186490 |
| 8 Aug |  | 52984 | 45635 | 61124 |  | 1983629 | 1748632 | 2247614 |
| 9 Aug |  | 54089 | 46629 | 62347 |  | 2037718 | 1795261 | 2309961 |
| 10 Aug |  | 55210 | 47637 | 63587 |  | 2092928 | 1842898 | 2373548 |
| 11 Aug |  | 56347 | 48661 | 64845 |  | 2149275 | 1891559 | 2438393 |
| 12 Aug |  | 57500 | 49700 | 66120 |  | 2206775 | 1941259 | 2504513 |
| 13 Aug |  | 58670 | 50755 | 67412 |  | 2265445 | 1992014 | 2571925 |
| 14 Aug |  | 59857 | 51825 | 68722 |  | 2325302 | 2043839 | 2640647 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 15 Aug | | 61060 | 52910 | 70050 | | 2386362 | 2096749 | 2710697 |
| 16 Aug | | 62280 | 54012 | 71396 | | 2448642 | 2150761 | 2782093 |
| 17 Aug | | 63517 | 55129 | 72759 | | 2512159 | 2205890 | 2854852 |
| 18 Aug | | 64771 | 56263 | 74141 | | 2576930 | 2262153 | 2928993 |
| 19 Aug | | 66042 | 57412 | 75541 | | 2642972 | 2319565 | 3004534 |
| 20 Aug | | 67330 | 58578 | 76959 | | 2710302 | 2378143 | 3081493 |
| 21 Aug | | 68637 | 59761 | 78396 | | 2778939 | 2437904 | 3159889 |
| 22 Aug | | 69960 | 60960 | 79852 | | 2848899 | 2498864 | 3239741 |
| 23 Aug | | 71302 | 62176 | 81326 | | 2920201 | 2561040 | 3321067 |
| 24 Aug | | 72661 | 63408 | 82820 | | 2992862 | 2624448 | 3403887 |
| 25 Aug | | 74039 | 64658 | 84333 | | 3066901 | 2689106 | 3488220 |
| 26 Aug | | 75434 | 65925 | 85864 | | 3142335 | 2755031 | 3574084 |
| 27 Aug | | 76848 | 67209 | 87416 | | 3219183 | 2822240 | 3661500 |
| 28 Aug | | 78280 | 68510 | 88986 | | 3297463 | 2890750 | 3750486 |
| 29 Aug | | 79731 | 69829 | 90577 | | 3377194 | 2960579 | 3841063 |
| 30 Aug | | 81201 | 71166 | 92187 | | 3458395 | 3031745 | 3933250 |
| 31 Aug | | 82690 | 72520 | 93817 | | 3541085 | 3104265 | 4027067 |

LPL: Lower Prediction Limit ; UPL: Upper Prediction Limit

## 4. Time Series Modelling of Daily +ve

Intuitively the number of positive cases detected on a particular day is very much likely to depend on the number of cases found on previous days. Therefore, we would like to model the Daily +ve data using time series modelling. In particular we shall be fitting ARIMA model to the data. In order to do this, we need to plot the auto-correlation function (ACF) and partial auto-correlation function (PACF) of the data on Daily +ve. The ACF and PACF of the Daily +ve data is shown in the figure 4 and figure 5 respectively. The tapering spikes of ACF as the lag increases suggests an AR model and the significance of PACF only at lag 1 indicates that an AR(1) model could appropriate. Before fitting of the model, we need to make sure that the data is stationary but we know that there is an exponentially increasing trend in the data. Hence, data is not stationary and we need to make the transformation.

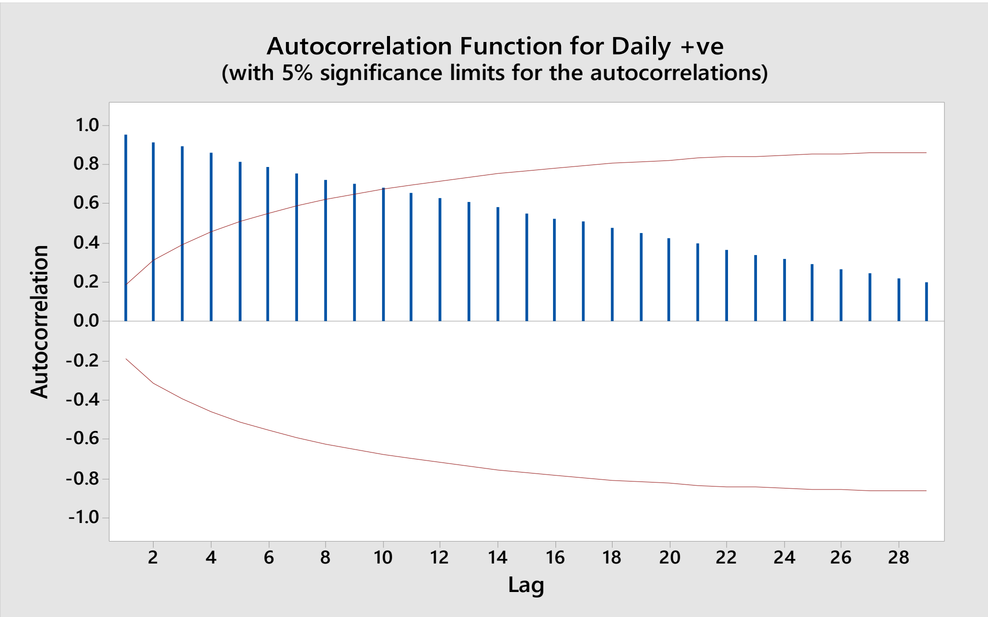


*Figure 4: ACF of Daily +ve*

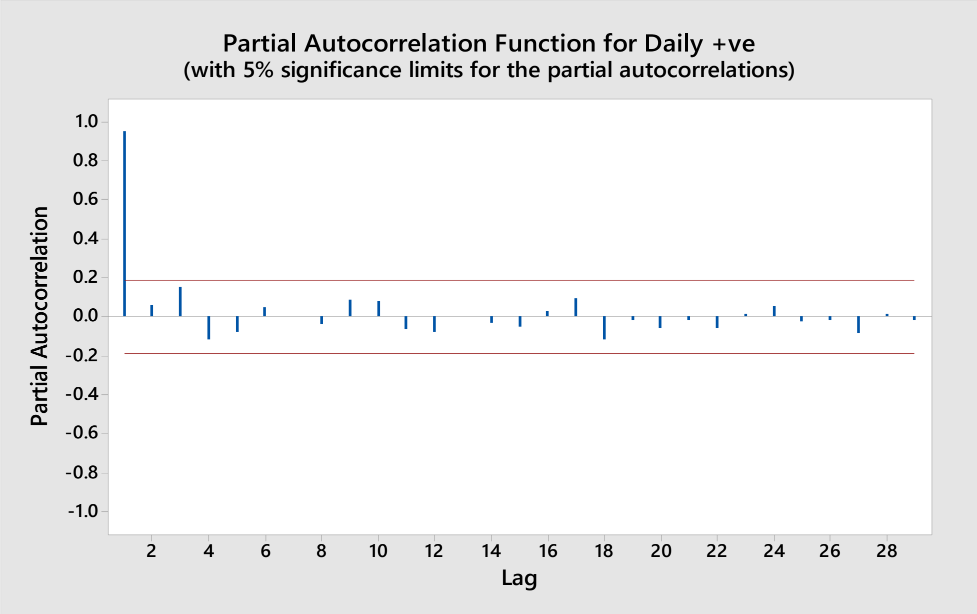


*Figure 5: PACF of Daily +ve*

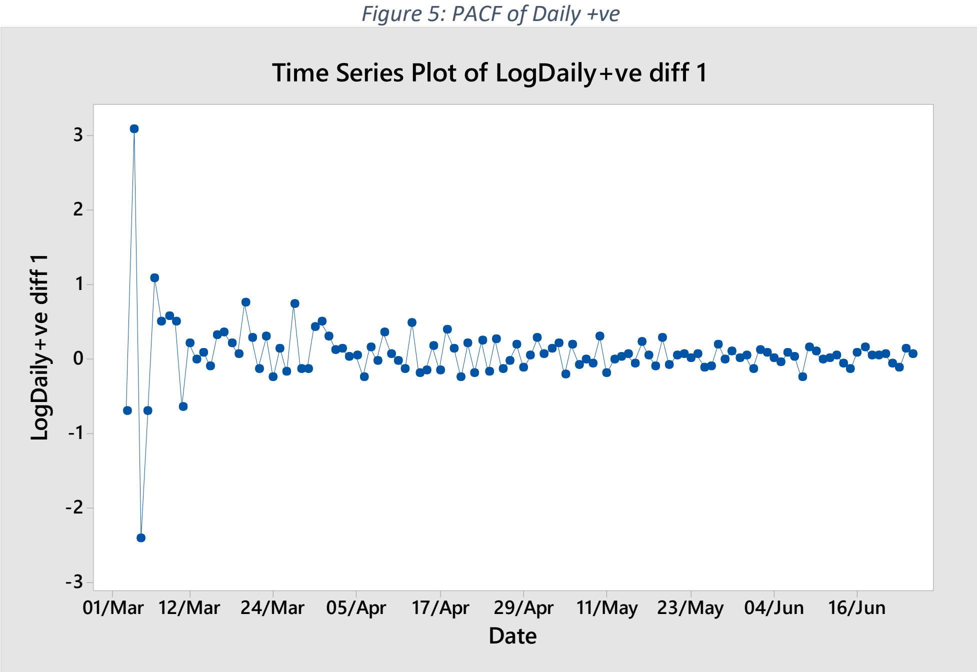


*Figure 6: Time series plot of first order differenced log Daily +ve*

We first filtered the data using first order differencing yet there was too much fluctuation in the variation and the stationarity could not be achieved. To reduce the variation, we tried the applying log on the original data and then finding the first order differencing. The time series plot of this filtered data is shown in the Figure 6. If we inspect the figure 6, we can observe that except for few points in the beginning the plot looks stationary. Therefore, we decided to drop the data for 2$^{nd}$ March through 11$^{th}$ March and plot the first order differenced log (Daily +ve) since 12$^{th}$ March onwards up to 24$^{th}$ June, the plot of which for 105 observations is given in Figure 7. It can be noticed that if it weren't for the few points like 20$^{th}$, 27$^{th}$,30$^{th}$ and 31$^{st}$ March the plot in Figure 7 is approximately stationary. Similar is the case with second order differenced data of log(Daily +ve) for 12$^{th}$ March onwards. Its plot is produced as the figure 8.

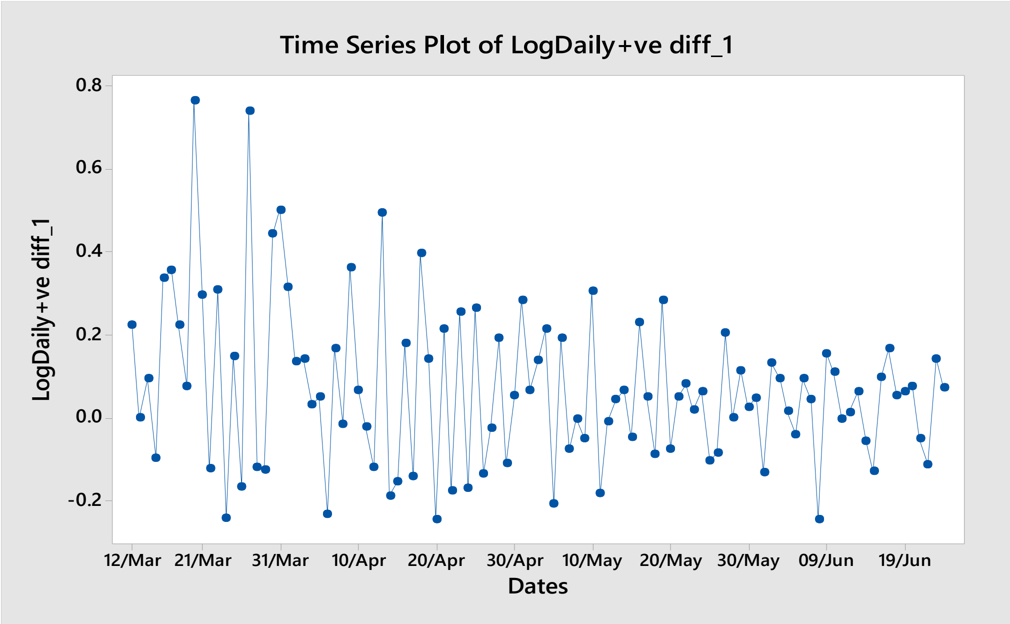


*Figure 7: Time series plot of first order differenced log Daily +ve for the period 12th March – 24th June*

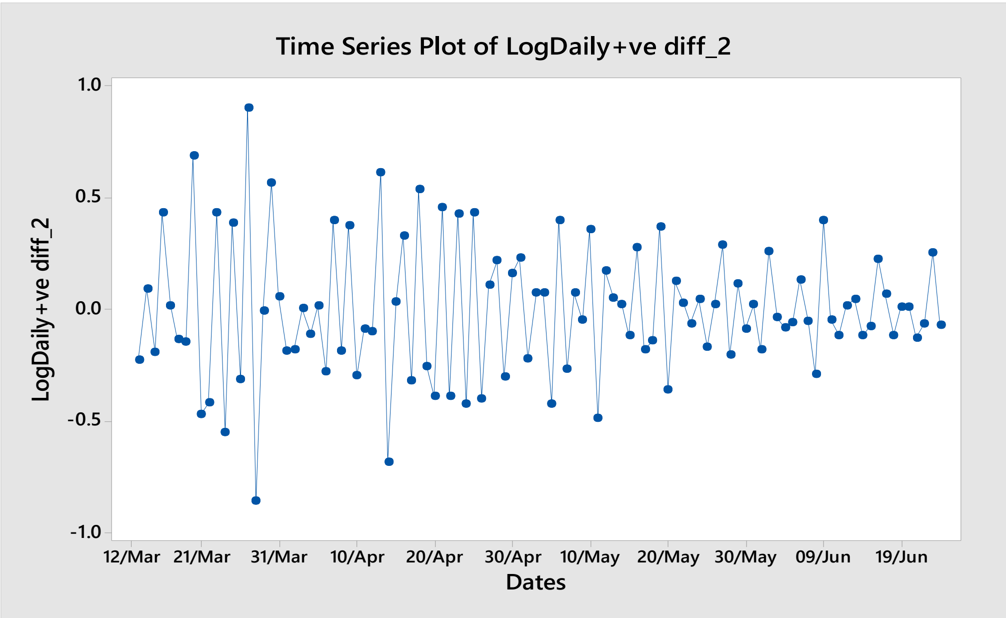


*Figure 8: Time series plot of second order differenced log Daily +ve for the period 12th March – 24th June*

Hence we fit several different models for log(Daily +ve). The best fit model we found is ARIMA(2,1,1) with no constant. The final estimates of the parameters are given in the Table 4. The significance of the parameters is tested using t-test with p-value very small for all. This indicates that all the parameters are significant. The model fit is

$$\mathbf{y_t = 0.7730\, y_{t-1} + 0.2247\, y_{t-2} + e_t + 0.95\, e_{t-1}} \qquad (2)$$

where $y_t = x_t - x_{t-1}$ and $x_t$ = log(Daily +ve) at time t. We can express the model in (2) in terms of $x_t$ by simply substituting for $y_t$.

*Table 4: Final estimate of the parameters*

| Type | Coef | SE Coef | T-Value | P-Value |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| **AR 1** | 0.7730 | 0.0994 | 7.77 | 0.000 |
| **AR 2** | 0.2247 | 0.0980 | 2.29 | 0.024 |
| **MA 1** | 0.9500 | 0.0330 | 28.83 | 0.000 |

The table 5 displays the Mean Sum of Square which is smallest for (2) as compared to the other models.

*Table 5: Residual sum of squares*

| DF | SS | MS |
|---|---|---|
| 101 | 3.57837 | 0.0354295 |

The residuals must not be auto correlated and this is apparent in Figure 9 with autocorrelation being insignificant at all the lags.

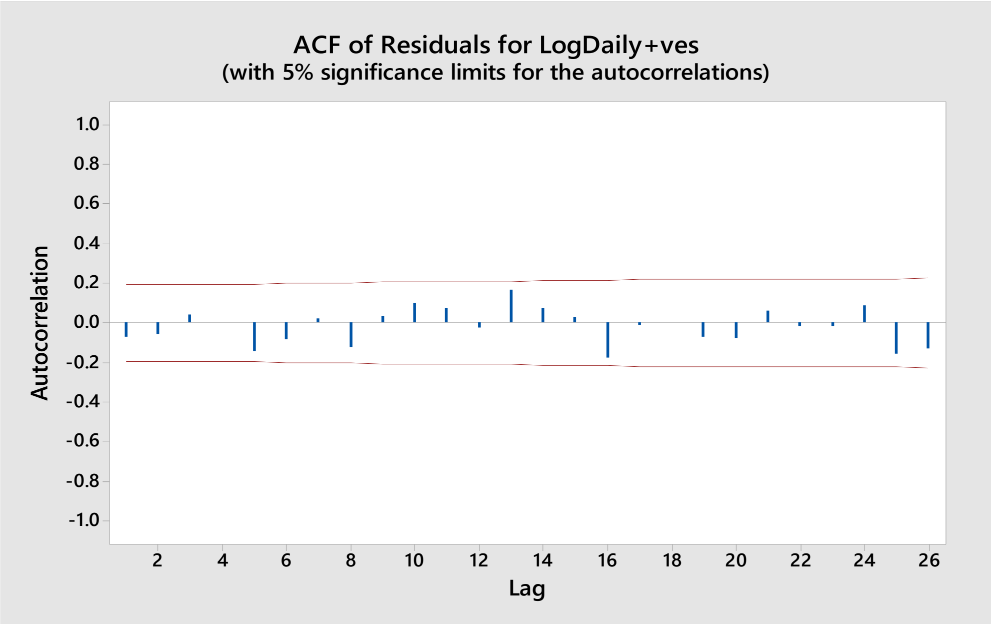


*Figure 9: ACF of residuals for ARIMA(2,1,1)*

The significance of the autocorrelation among residuals at four lags is tested using Ljung-Box Chi-Square Statistics and output is shown in the Table 6. The p-values for the test at all the four lags are very large pointing towards the insignificance of autocorrelation among residuals.

*Table 6: Modified Box-Pierce (Ljung-Box) Chi-Square Statistic*

| **Lag** | 12 | 24 | 36 | 48 |
|---|---|---|---|---|
| **Chi-Square** | 8.02 | 19.14 | 30.78 | 37.65 |
| **DF** | 9 | 21 | 33 | 45 |
| **P-Value** | 0.533 | 0.576 | 0.578 | 0.773 |

Also, the residuals are normally distributed which can be observed in the NPP and histogram in top-left and bottom-left corner of the Figure 10. The error has constant variance which is evident from the graph of residual against the fitted values in the top-right corner of the figure 10. One may have the

impression that the variance is not constant for the error but it can be noticed that it is happening due to the just three residual values at the top of the graph out of 101 residuals which can be ignored.

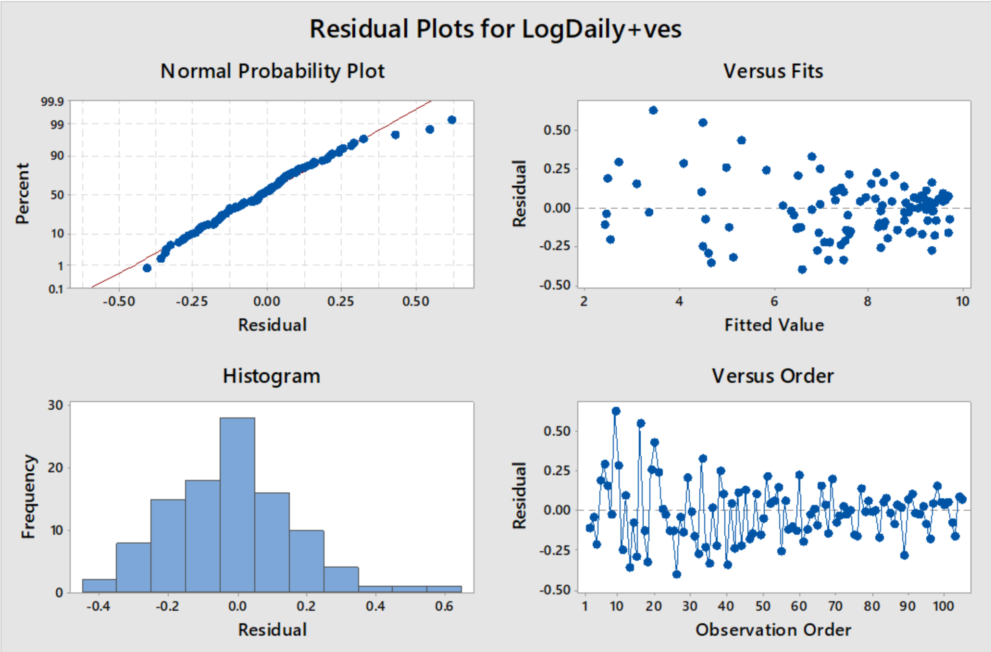


*Figure 10: residual plot for ARIMA(2,1,1)*

Now that we have fit the best possible model with all the assumptions being satisfied, we move to our main purpose of forecasting. But before that we would like to see how well the model fits in the data graphically. This can be seen in the figure 11 which is time series plot of the observed and fitted Daily +ve.

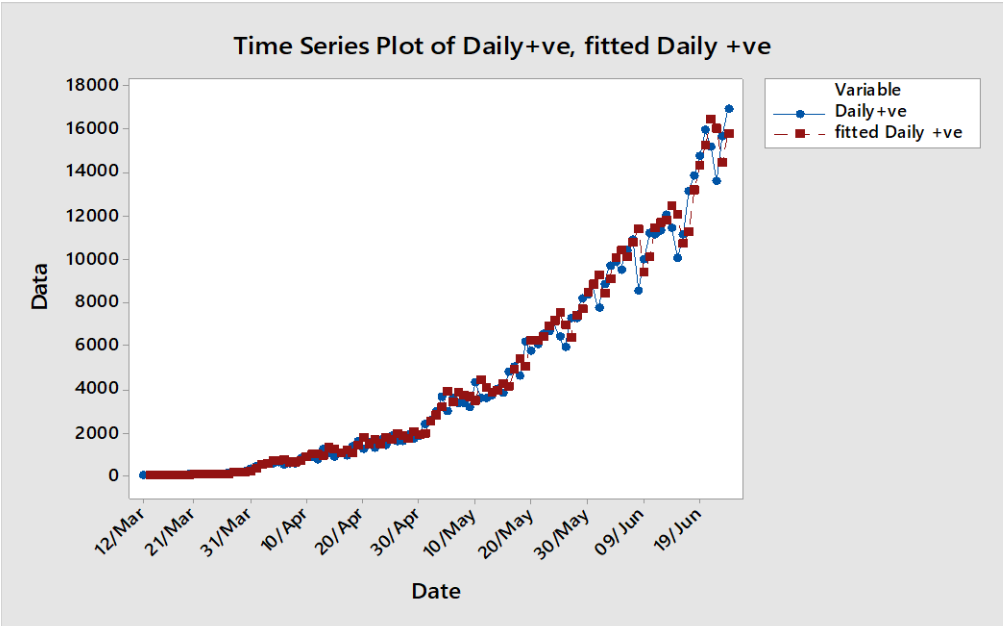


*Figure 11: Time series plot of observed Daily +ve and fitted Daily +ve through ARIMA(2,1,1)*

The forecast for lead time = 10, i.e. for 10 days ahead of 24$^{th}$ June for log(Daily +ve) is given in the table 7 with 95% confidence limits.. We must undo the transformation to get the forecasts for Daily +ve. For this we need to raise the given forecasts and the confidence limits to the power of e.

*Table 7: Forecast for 10 Days from 25th June for Log(Daily +ve) using ARIMA(2,1,1)*

| | | | 95% Limits | |
|---|---|---|---|---|
| **Period** | **Date** | **Forecast** | **Lower** | **Upper** |
| 106 | 25 June | 9.7593 | 9.39033 | 10.1283 |
| 107 | 26 June | 9.7963 | 9.31839 | 10.2742 |
| 108 | 27 June | 9.8308 | 9.24647 | 10.4151 |
| 109 | 28 June | 9.8657 | 9.18636 | 10.5451 |
| 110 | 29 June | 9.9005 | 9.13073 | 10.6702 |
| 111 | 30 June | 9.9352 | 9.07848 | 10.7919 |
| 112 | 1 July | 9.9699 | 9.02831 | 10.9114 |
| 113 | 2 July | 10.0045 | 8.97954 | 11.0294 |
| 114 | 3 July | 10.0390 | 8.93165 | 11.1463 |
| 115 | 4 July | 10.0734 | 8.88428 | 11.2626 |

The back-transformed forecasts are given in the Table 8 rounded to the nearest integers.

*Table 8: Forecasts for 10 Days from 25th June for Daily +ve using ARIMA(2,1,1)*

| | | | 95% Limits | |
|---|---|---|---|---|
| **Period** | **Date** | **Forecast** | **Lower** | **Upper** |
| 106 | 25 June | 17315 | 11972 | 25042 |
| 107 | 26 June | 17967 | 11141 | 28975 |
| 108 | 27 June | 18598 | 10368 | 33360 |
| 109 | 28 June | 19258 | 9763 | 37991 |
| 110 | 29 June | 19940 | 9235 | 43054 |
| 111 | 30 June | 20644 | 8765 | 48625 |
| 112 | 1 July | 21373 | 8336 | 54798 |
| 113 | 2 July | 22126 | 7939 | 61661 |
| 114 | 3 July | 22902 | 7568 | 69307 |
| 115 | 4 July | 23704 | 7218 | 77855 |

## 5. Analysis of Active Cases

Define $X_t$ = Total Active Cases on day t

Define $C_t = X_{t+1}/X_t$

This is the ratio of total active cases today and same on yesterday. As long as the total active cases keep increasing the ratio $C_t$ will be greater than unity. But when the total active cases today are greater than yesterday then this ratio falls below unity and remain consistently below but close to unity if the

situation keeps getting better every day. Therefore, $C_t$ can be used to predict the day on and after which the active cases start rolling down. This will happen on the day $C_t$ falls below unity.

The ratio $C_t$ cannot be used to predict the end of the pandemic. This ratio will become zero only if the for a given t, $X_{t+1}$ is zero but this does not indicate the eradication of the pandemic because the on t+2 active cases may appear again. Besides technically $C_t$ can be equal to unity if on two or more consecutive days the number of new active cases are same even though these cases are large in number.

To understand how this ratio $C_t$ has been varying in India, see the figure 12 which plots $C_t$ against t from 4th April through 16th July.

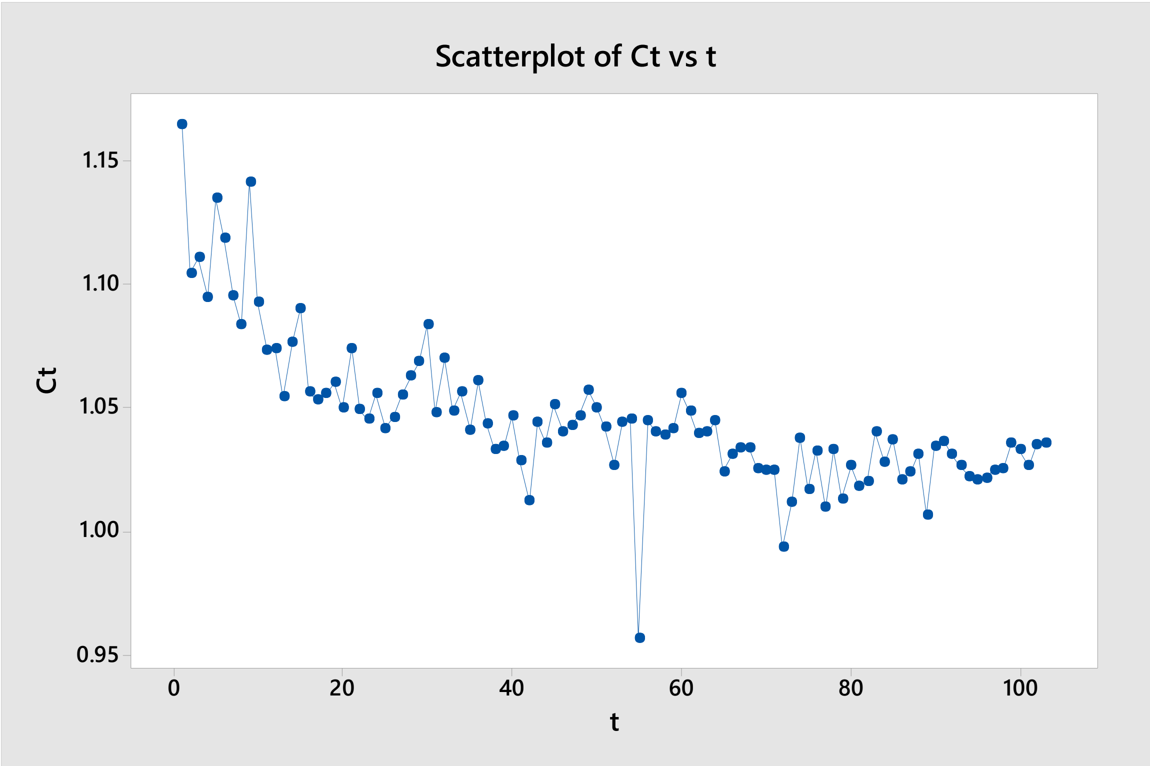


*Figure 12*

We can observe that this ratio $C_t$ has been decreasing exponentially. To linearize the relationship, we plotted the graph of $\log C_t$ against t and fitted the model but it wasn't satisfactory. We also plotted the $\log C_t$ and logt. The plot is given below in figure 13 and we can observe the linear relationship between $\mathrm{Log} C_t$ and logt with negative slope.

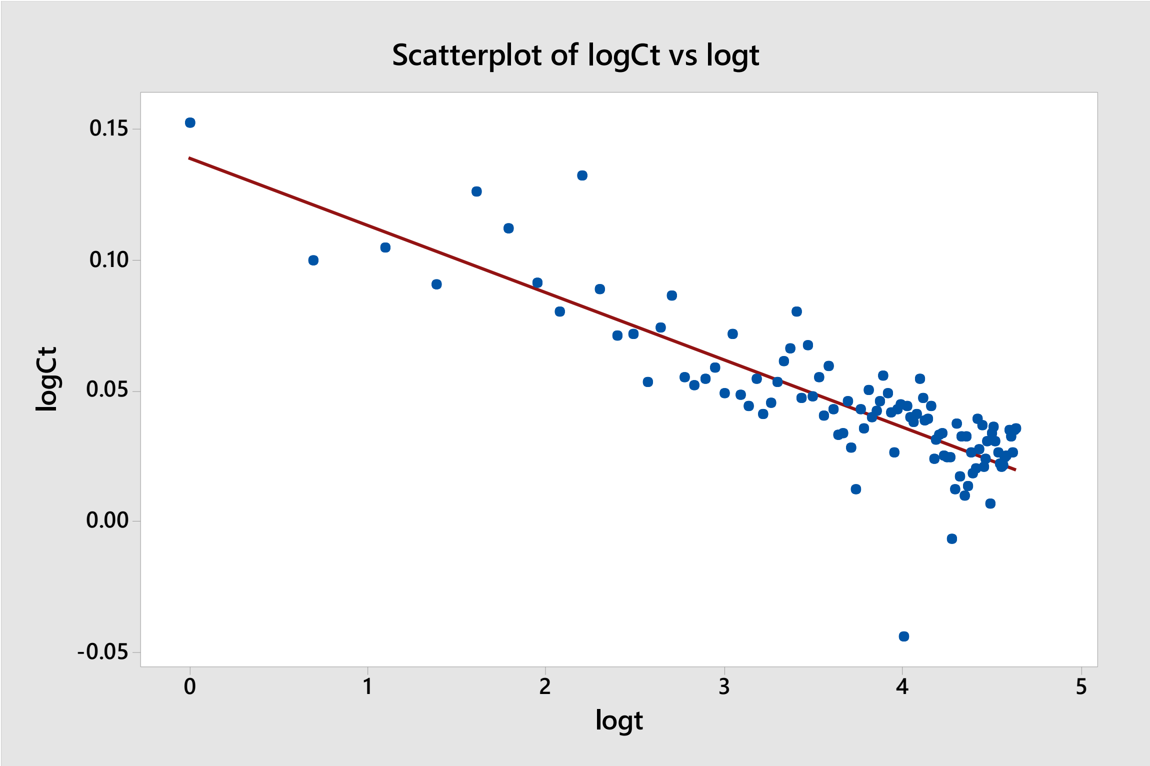


*Figure 13*

We fit the linear regression for $logC_t$ against logt

# Regression Analysis: $logC_t$ versus logt

## Regression Equation

$$logC_t = 0.13961 - 0.02578 \, logt \quad (3)$$

## Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 1 | 0.05857 | 0.058565 | 261.82 | 0.000 |
| logt | 1 | 0.05857 | 0.058565 | 261.82 | 0.000 |
| Error | 101 | 0.02259 | 0.000224 | | |
| Total | 102 | 0.08116 | | | |

## Model Summary

| R-sq | R-sq(pred) |
|---|---|
| 72.16% | 70.89% |

## Coefficients

| Term | Coef | SE Coef | T-Value | P-Value |
|---|---|---|---|---|
| Constant | 0.13961 | 0.00602 | 23.17 | 0.000 |
| logt | -0.02578 | 0.00159 | -16.18 | 0.000 |

## Durbin-Watson Statistic

Durbin-Watson Statistic = 1.82298

We can observe Durbin-Watson statistic is close to 2 indicating no autocorrelation among the residuals. Also, the residual plot shown in figure 14 justifies the validity of assumptions.

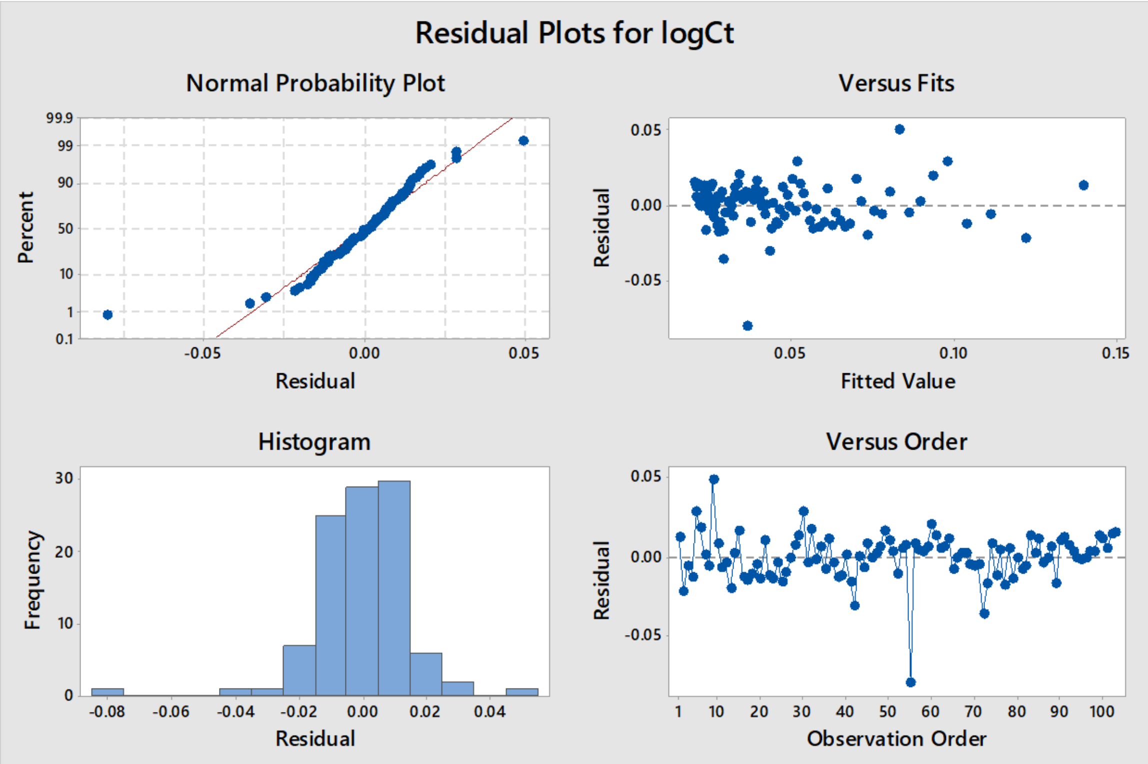


Figure 14

The figure 15 plots the $R_t$ and fitted $R_t$ found using the regression model **logRt = 0.13961 - 0.02578 logt**

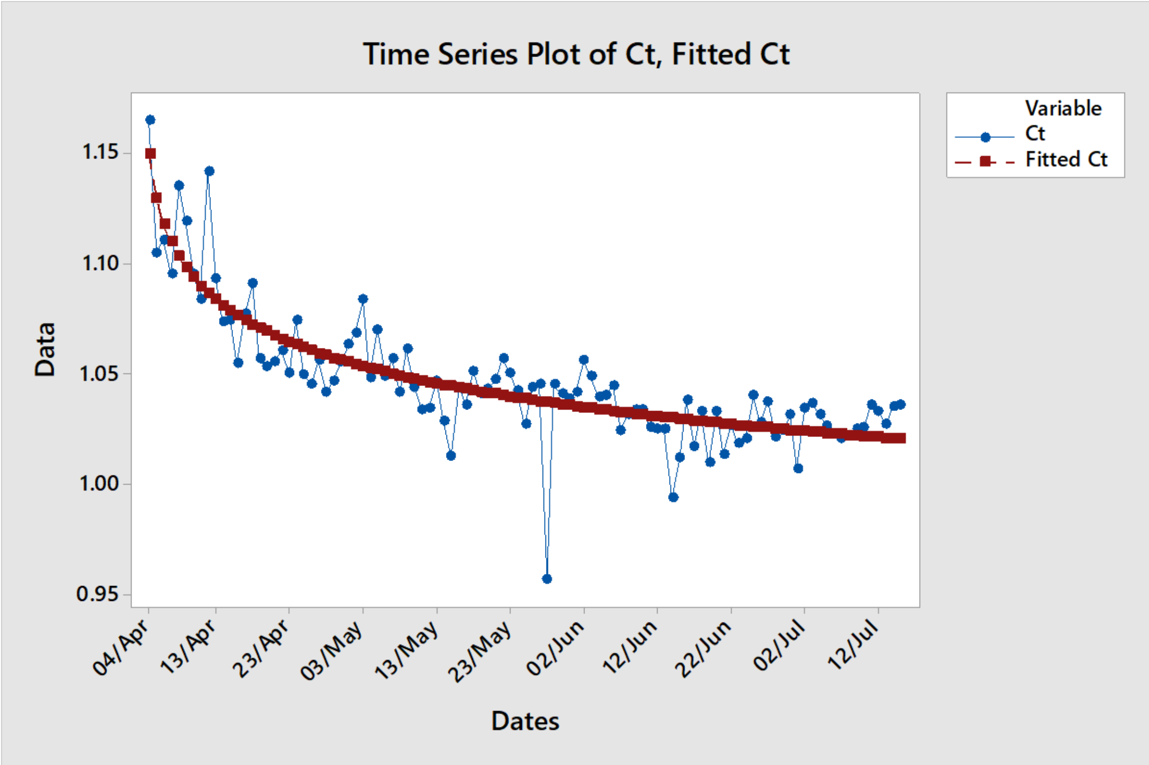


*Figure 15*

Now we want to determine the day on which ratio $R_t$ becomes 1. That is, the day Total Active Cases ($X_t$) starts rolling down. For this we solve the regression equation for t when $R_t = 1$ and we find that for t = 229, $R_t = 1$. In our data t = 1 for 4th April. Hence, as per the model it can be expected that from about 18th Nov 2020 the new Active Cases found daily starts decreasing in India.

## 6. Analysis of Recovery Rate

Following is the graph of 'Recovery Rate' in India in figure 16.

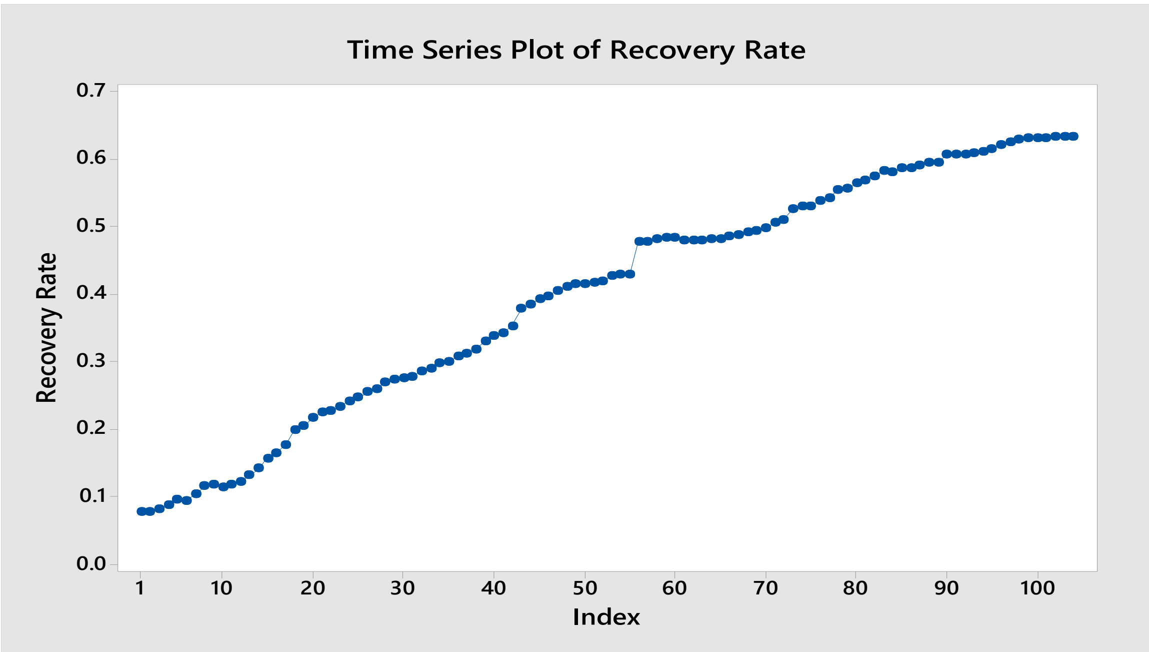


*Figure 16*

We define recovery rate = $R_t = D_t / Y_t$

Where $D_t$ = Total discharged/recovered patient until day t

$Y_t$ = total covied-19 positive cases until day t

We can observe that recovery rate has been consistently increasing. This means the rate of increment of recovering patients is greater than the rate of increment in total positive. The day recovery rate unity indicates on the day total number of recovered patients are equal to total number of positive cases. That is, all the positive cases found so far have been recovered and implies the end of the pandemic. Also, once this recovery rate hits unity any newly found cases will also be cured with the same rate and surge in the cases will be curbed.

$R_t$'s are undoubtedly autocorrelated and figure 17 shows the autocorrelation function of the $R_t$ for various lags. The tapering spikes indicates the model could be Autoregressive model.

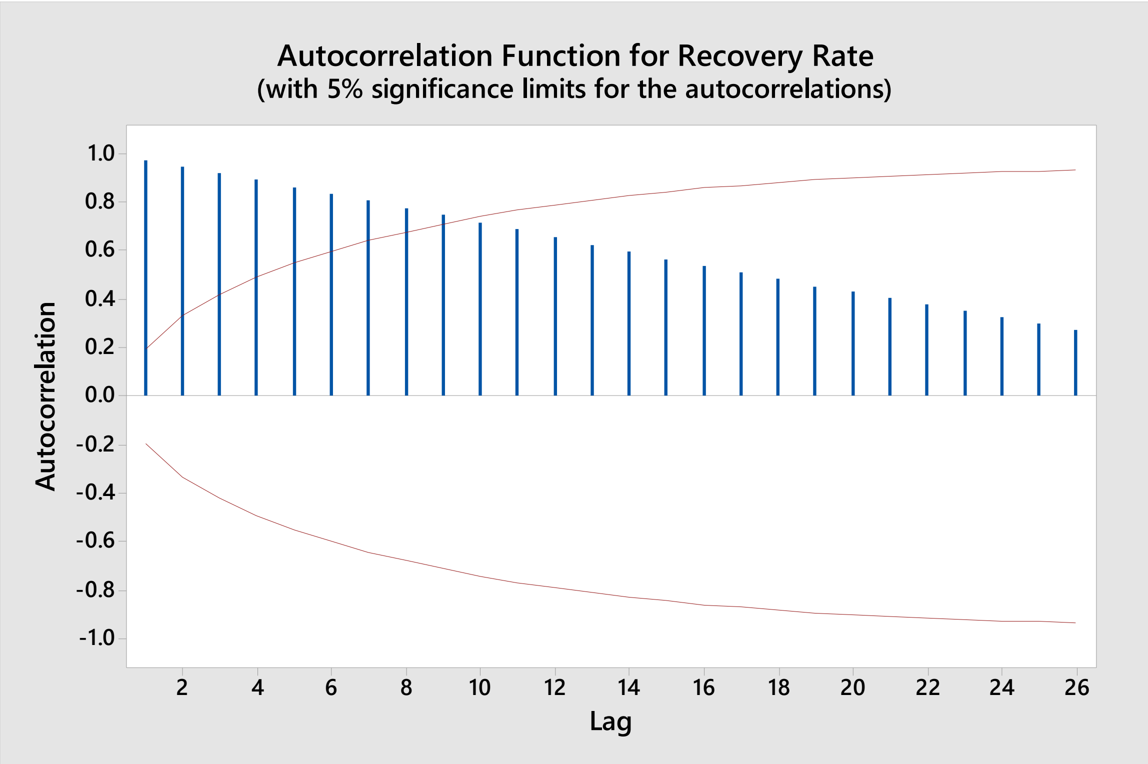


*Figure 17*

In the figure 18 we see that partial autocorrelation is significant only at lag 1. This indicates that the tentative model could be AR(1).

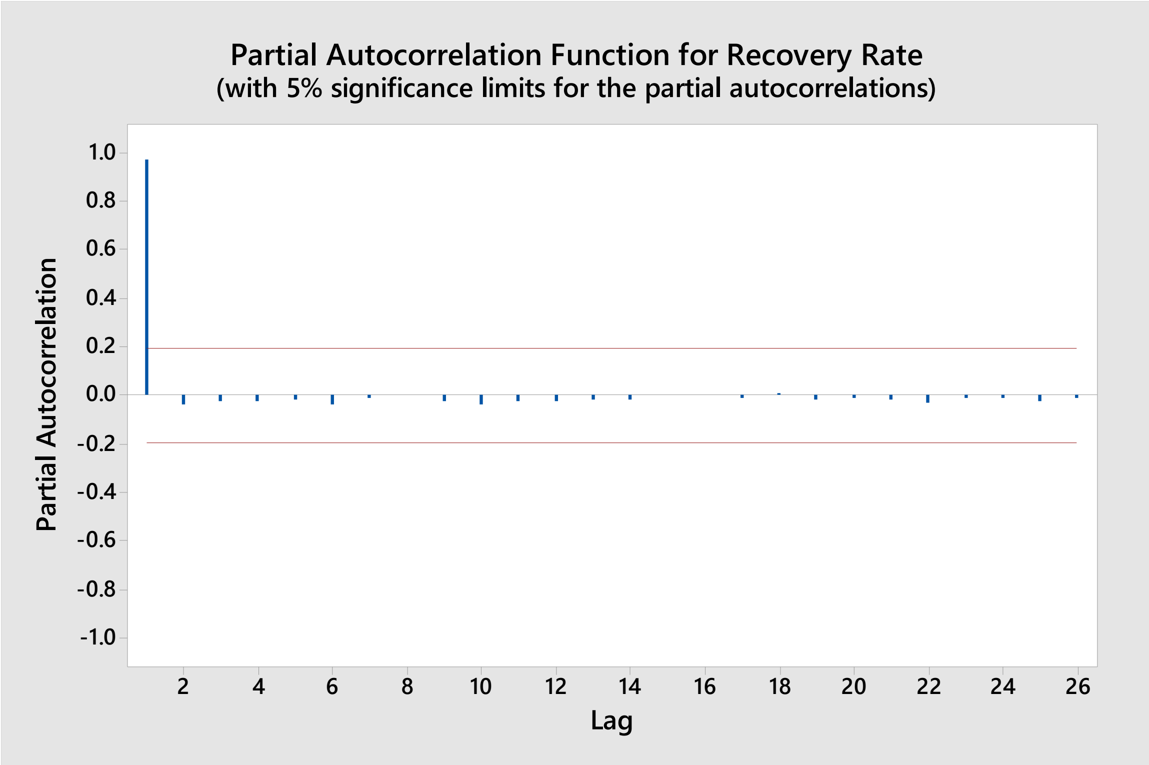


*Figure 18*

We model $R_t$ against $R_{t-1}$ using least square method.

# Regression Analysis: Recovery Rate versus Recovery Rate lag1

## Method

| | |
|---|---|
| Box-Cox transformation | |
| Rounded λ | 1 |
| Estimated λ | 1.02717 |
| 95% CI for λ | (0.997674, 1.05767) |

## Regression Equation

$$R_t = 0.00867 + 0.99170\, R_{t-1} \quad (4)$$

## Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 1 | 3.04794 | 3.04794 | 79146.34 | 0.000 |
| $R_{t-1}$ | 1 | 3.04794 | 3.04794 | 79146.34 | 0.000 |
| Error | 101 | 0.00389 | 0.00004 | | |
| Total | 102 | 3.05183 | | | |

## Model Summary

| R-sq | R-sq(pred) |
|---|---|
| 99.87% | 99.87% |

## Coefficients

| Term | Coef | SE Coef | T-Value | P-Value |
|---|---|---|---|---|
| Constant | 0.00867 | 0.00152 | 5.70 | 0.000 |
| $R_{t-1}$ | 0.99170 | 0.00353 | 281.33 | 0.000 |

## Durbin-Watson Statistic

Durbin-Watson Statistic = 1.95283

Durbin-Watson statistic is very close to 2 indicating here is not autocorrelation among residuals. Also, the residuals are satisfactorily justifying the assumption made for regression model.

### Residual Plots for Recovery Rate

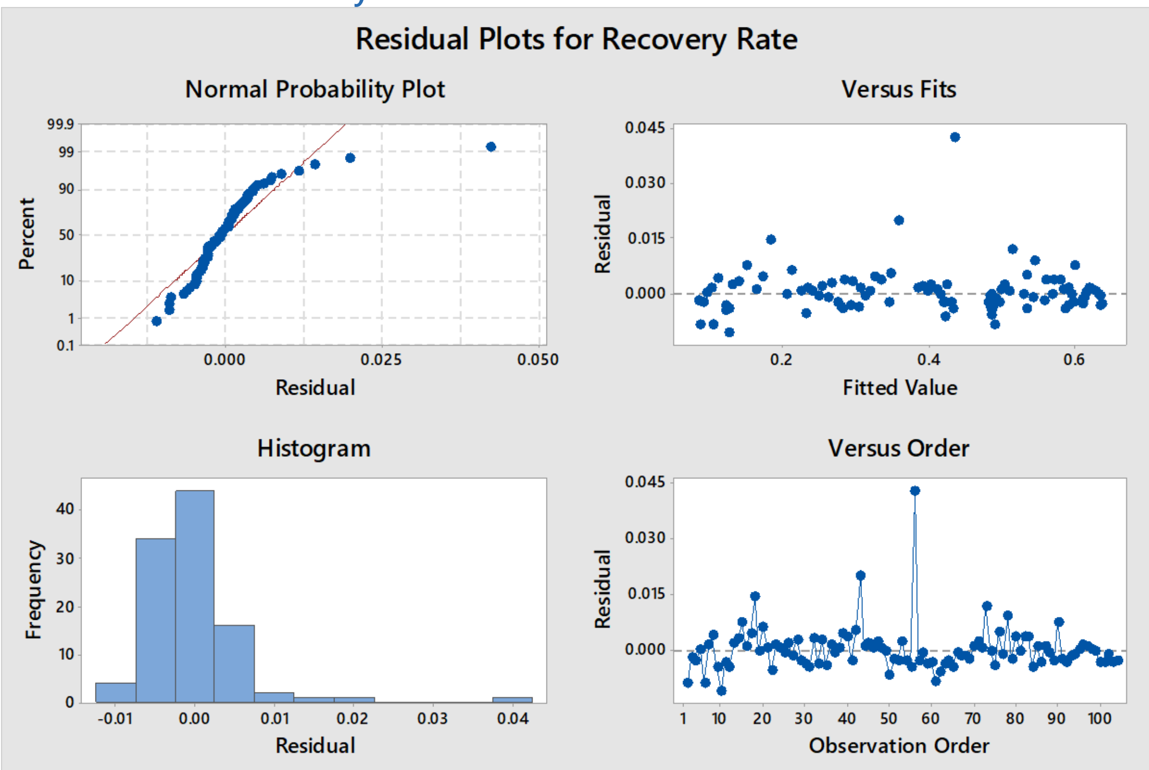


*Figure 19*

The figure 20 demonstrates the observed and fitted recovery rate. Now we will be determining the 't' for which $R_t$ hits unity. So, if the situation remains the same then as per above model, the recovery rate will become one on 368th day where 4th April 2020 is the day one. This means by 9th April 2021 the recovery rate reaches 100%. This indicates all the covid-19 patients will be recovered by about 9th April 2021 and hence the pandemic will end in India.

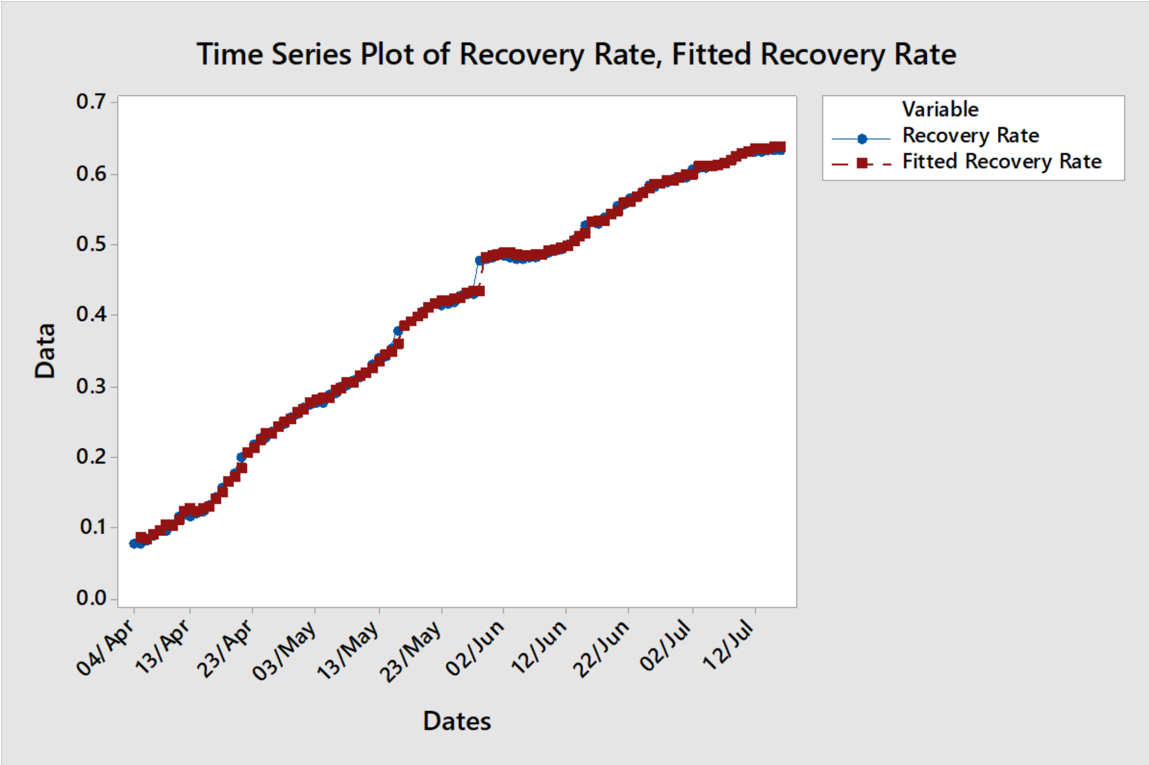


*Figure 20*

It must be mentioned that in case of invention of effective vaccines the recovery rate could improve even more and the pandemic can end even earlier. Figure 21 shows the observed and predicted recovery rate.

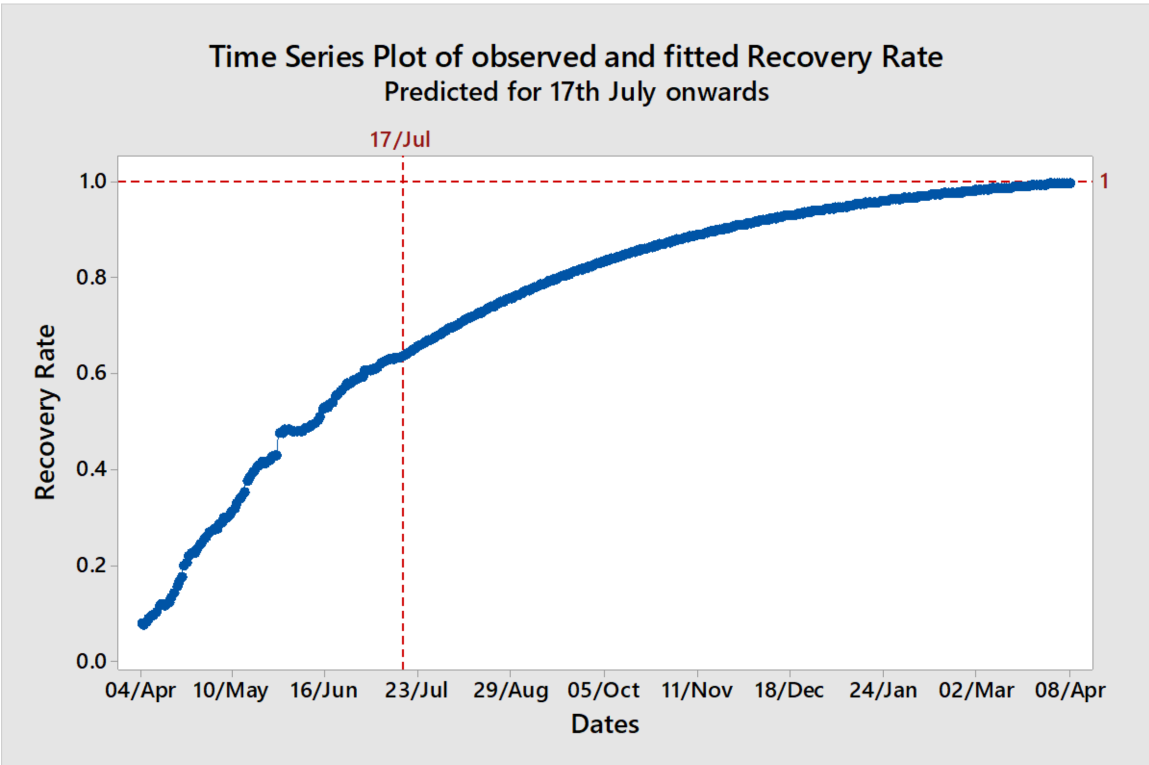


*Figure 21*

Following table lists some predicted recovery rate in percentage for various dates.

*Table 9*

| Date | Recovery rate % | Date | Recovery rate% |
|---|---|---|---|
| 16/07/2020 | 63.3 | 31/10/2020 | 87.6 |
| 21/07/2020 | 65.0 | 19/11/2020 | 90.1 |
| 31/07/2020 | 68.1 | 30/11/2020 | 91.3 |
| 07/08/2020 | 70.2 | 31/12/2020 | 94.3 |
| 25/08/2020 | 75.0 | 08/01/2021 | 95.0 |
| 31/08/2020 | 76.4 | 31/01/2021 | 96.6 |
| 17/09/2020 | 80.1 | 28/02/2021 | 98.3 |
| 30/09/2020 | 82.6 | 31/03/2021 | 99.7 |
| 14/10/2020 | 85.0 | 08/04/2021 | 100 |

The following R-command is used to predict the recovery rate ($R_t$) iteratively.

```
R=NULL

R[1]=0.633031

for(i in 2:270){

  RR[i]=0.00867+0.9917*RR[i-1]

}

R
```

## 7. Conclusion

1) In India, the model predicted that the cumulative number of positive cases may hit the figure of 10,00,000 by 16th of July and it happened exactly on the same day with 10,05,637 cases. (refer Table 3).
2) By the end of July, the cumulative +ve cases may rise to 15 lacs as per the model. The 20 lac cumulative cases may be hit by the end of first week of August. By the end of Aug the total cases may rise beyond 35 lacs.
3) The situation remains the same overall for the whole country and the Daily +ve cases rise at the same rate they have been rising so far then from third week of July every fourth day 1 lac new cases will be added to the number of cumulative +ve cases. (refer Table 3).
4) As per the model the Daily +ve cases can hit the mark of 50,000 by 6th Aug. It may happen by 30th July at the earliest as per the model.
5) As per the model (3) it can be expected that from 18th Nov 2020 the new Active Cases added daily starts decreasing in India.
6) **As per model (4) all the covid-19 patients will be recovered by second week of April 2021 and hence the pandemic will end in India provided the conditions remain roughly the same.**

---

Dr. Ulhas J Dixit (ulhasdixit@yahoo.co.in)

Priyesh D Tiwari (tiwaripds@gmail.com)